International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 6 Issue 3 PP 123-128, 2025
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V6I3P114

Original Article

Cybersecurity Through AI-Powered, Distributed Intrusion Detection And Prevention Systems

Naresh Kalimuthu
Independent Researcher.

Received On: 08/07/2025 ***Revised On: 27/07/2025*** ***Accepted On: 28/08/2025*** ***Published On: 10/09/2025***

Abstract *- The growing sophistication of zero-day attacks has rendered traditional Intrusion Detection and Prevention Systems (IDPS) almost ineffective in enterprise networks. In this paper, we explore the transition to AI-based distributed IDPS, focusing particularly on Federated Learning (FL) as a core architecture. This approach provides enhanced, adaptive threat detection with built-in privacy protections. However, implementing this method in practice presents several challenges. This work addresses three key issues: the balance between scalability and computational overhead, privacy concerns in FL, and the vulnerability of AI to adversarial attacks. We incorporate cutting-edge solutions and draw on real-world examples to argue that only a multi-layered strategy combining architectural, cryptographic, and model-hardening measures can fully unlock the potential of these next-generation security systems.*

Keywords *- Intrusion Detection System (IDS), Artificial Intelligence (AI), Machine Learning, Federated Learning (FL), Distributed Systems, Cybersecurity, Anomaly Detection, Adversarial Attacks.*

1. Introduction

The growing digital environment faces an uneven battle between attackers and defenders in cyberspace. Every day, attackers develop new and sophisticated ways to bypass security measures, revealing the shortcomings and obsolescence of traditional intrusion detection and prevention systems (IDPS). The main flaw of IDPS is their reliance on “signature systems” for intrusion detection. This method of network traffic analysis, known as 'signature detection,” is fundamentally outdated, reactive, and ineffective. Additionally, detecting zero-day attacks is nearly impossible. The system would need to constantly and exhaustively target a signature database to be somewhat effective against current threats. Contrary to popular belief, unbounded detection systems capable of identifying anything (anomaly-based detection systems) would only keep systems in a ‘safe mode.’ Their practical use has been limited due to overwhelming the systems with high false positives.

The inadequacy of a signature-based approach to intrusion detection systems has sparked an AI (artificial intelligence) revolution in cybersecurity. The main shift being introduced is the move from static rule-based monitoring of the quest systems to dynamic, adaptive, and responsive defense. These systems are adaptable and accurate when predicting incoming attacks and learning new ones, thanks to their low false alarm rates. Support Vector Machines (SVMs), Random Forests, Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks have performed exceptionally well on complex tasks crucial for intrusion detection, reinforcing sophisticated neural network techniques. Data privacy and protection are essential pillars in today's digital world. The centralized system of training IDPS systems, which aggregates data from the entire network to a single point, significantly increases latency and creates a single point of failure with the bounded model. This model fails to handle the overwhelming data in IoT and enterprise cloud systems or scale to meet exponential growth. Unfortunately, it also falls short in offering the necessary data privacy and protection measures that are so important today. These requirements can seem to conflict with company laws and policies. To address these architectural hurdles, we moved towards distributed IDPS frameworks.

Within this framework, Federated Learning (FL) has emerged as the most promising and impactful advancement. FL is a type of distributed system where a global model is trained collaboratively and in parallel by multiple clients (e.g., edge devices, servers) without exchanging their raw, privacy-sensitive data. Instead, clients perform local training and only send a few model updates, such as gradients, to the central server for aggregation. This data-centric approach preserves privacy by design, reduces the communication bandwidth needed to transmit the complete dataset, and enhances overall scalability. This paper examines the architecture, challenges, and solutions for AI-powered, distributed IDPS, especially on the Federated Learning (FL) framework. It begins by identifying three core research challenges within the IDPS architecture: inherent privacy and security vulnerabilities, the misuse of AI models, and vulnerabilities in federated systems architecture that can lead to adversarial AI, along with the systemic scalability issue. It then presents a set of countermeasures to address these challenges. The document proceeds by integrating these countermeasures with real-world case studies to demonstrate the effectiveness of such advanced systems. Finally, it highlights emerging trends in this vital area of cybersecurity.

2. Research Topics: Core Challenges in AI-Powered Distributed IDPS

Deploying AI-powered, distributed IDPS successfully isn't just about implementing advanced algorithms in a decentralized setup. It involves navigating a complex balance of competing needs. Pursuing higher detection accuracy through more complex models often requires more computing power, which strains edge devices and impacts system scalability. Implementing strong privacy-preserving measures can also increase communication and computing overhead, potentially harming real-time performance. Additionally, making models more resilient to one type of threat, like adversarial evasion, might not protect the entire federated system from internal attacks. This creates a challenging optimization problem where security, privacy, and performance must be carefully balanced. The following sections explore three key challenges that define the main aspects of this problem.

2.1. Scalability, Latency, and Computational Overhead

One of the main reasons for adopting a distributed architecture is to avoid the performance limitations of centralized systems. However, decentralization brings its own performance challenges. Traditional cyber threat intelligence (CTI) and Intrusion Detection Systems (IDS) struggle to analyze data flows spanning multiple petabytes within IoT ecosystems and enterprise networks. This results in high data processing latency and limited scalability, which can be disastrous. While distributed systems eliminate the single-point bottleneck, they introduce new challenges related to system complexity, maintaining consistent data across nodes, and increased network latency due to the communication required between services for global system coherence.

Adding to the problem are the resource-hungry AI models themselves. While deep learning architectures are effective at detecting anomalies, they require substantial computational power during both training and real-time use. This makes it hard to deploy them beyond network edge devices like IoT sensors, mobile devices, and industrial controllers, which are all limited by energy, computing power, or memory. Additionally, while the Federated Learning (FL) framework has its benefits, it also brings new performance challenges. Although FL doesn't require transferring raw data, it's essentially an iterative process that involves multiple rounds of communication. In each round, the system model is sent to clients, who then return local updates. This back-and-forth can create too much communication overhead in low-bandwidth networks, resulting in slower convergence and higher latency.

2.2. Privacy and Security Risks in Federated Architectures

At its core, Federated Learning aims to protect data privacy. However, this creates a key paradox: while raw data remains on the local device, the model updates shared during training can still be used to compromise that very privacy. Although the gradients and model weights sent by each client are abstract, they still contain some implicit information about the local data used to generate them. This creates a new attack surface that could be exploited by a malicious central server or an untrustworthy participant in the federated network. A major privacy concern is inference attacks, where an attacker attempts to reconstruct an individual's private information by analyzing shared model updates. Research shows that such attacks can recover complete gradients and extract representative samples of the training data with high fidelity. In cybersecurity, this could lead to the compromise of sensitive configurations, proprietary communication systems, or private user actions.

Besides privacy breaches, it is essential to protect the integrity of the globally trained model from malicious users. These threats seek to compromise the model, making it unreliable or unsafe. Attacks typically fall into two categories:

- **Data Poisoning:** A malicious participant intentionally introduces incorrect, mislabeled, or corrupted data into the local training set. The resulting model update reflects these errors, which, when aggregated, degrade overall model performance or lead to misclassification of certain traffic types.
- **Model Poisoning:** This involves an attacker directly modifying the model update before sending it to the server. A particularly dangerous type is the backdoor attack, where the attacker embeds a hidden trigger in the model update. This allows the model to perform normally on most tasks but to misclassify certain inputs, such as mislabeling an adversarial dataset as benign when the trigger is present.

2.3. System Weaknesses: Attacks by Adversarial AI

Any defense mechanisms surrounding the architecture of learning systems are likely to be compromised. The AI detection systems, by their design, exhibit vulnerabilities to certain types of hostile manipulation. It is a well-known fact that AI systems, especially deep neural networks, are often more fragile than they are perceived to be. An adversary can insert some form of distortion into an input, and such distortion attacks of this sort can succeed without detection, often making them difficult to identify. By distributing the distortion in acceptable ways, the model is tricked into incorrect classification with high confidence. Such targeted intrusion strikes at the very core of AI's decision-making capacity.

Within an IDPS context, these attacks appear in several critical forms:

- **Evasion Attacks:** This type of attack is the most straightforward during system inference. An attacker takes malicious action, such as modifying a network packet containing an exploit. The goal of these modifications is to change the packet just enough that it crosses the model’s decision boundary and is classified as ‘benign traffic." Even though the payload is malicious, the attacker can get it past the system without triggering IDPS.

- **Poisoning Attacks:** It involves corrupting the very first, foundational training dataset. By injecting moderate, malicious samples into the dataset used to train the IDPS model before deployment, an attacker can impair the model's accuracy, creating extensive blind spots for specific attack types or backdoors that can be exploited later on.
- **Model Extraction and Inversion Attacks:** With the right skills, any attacker can do this if they have access to the deployed IDPS model. By systematically probing the IDPS system for its architecture or parameters, they can reverse engineer the system and obtain the model. This stolen model can then be used to create more advanced evasion attacks, making them more effective. In the case of model inversion attacks, privacy is also compromised since sensitive data can be reconstructed using the model's capabilities and outputs.

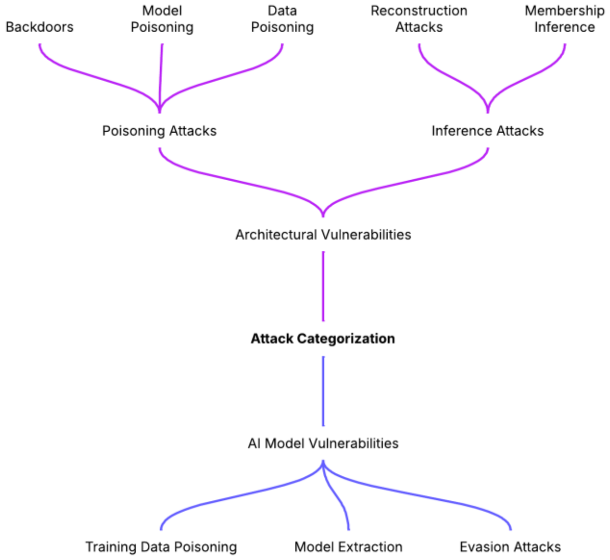

Fig 1: Security and Privacy Threats

3. Recommendations and Mitigation Strategies

Tackling the diverse and complex issues related to AI-powered distributed IDPS requires more than a single-point solution. A single, effective defensive strategy cannot rely on an advanced algorithm, a single secure protocol, or a strong architecture in isolation. Instead, a unified, multi-faceted security approach is necessary, integrating architectural design, privacy-preserving technologies, and model-level hardening. This layered strategy emphasizes a defensive system where an attacker must breach multiple diverse systems to succeed, making attacks more complex and costly. This forms the core of the resilient framework.

3.1. Architecting for Scalability and Efficiency

Distribution systems often encounter performance bottlenecks that can be effectively addressed using the Hybrid Edge-Server Framework. This approach shifts more complex tasks, such as the pre-training phase of large foundational models requiring significant computing power, to a powerful central server. Meanwhile, edge devices focus on fine-tuning the model. This pretraining method allows devices to access their local data without consuming heavy resources. It saves resources for device users and makes the system easier for sophisticated AI model users to operate. Complementing this architectural approach is the development of lightweight and optimized AI models. Instead of relying on heavy, congested resource networks, the focus should be on more efficient systems like convolutional neural networks.

These are designed to provide a better balance between loss and gain, resulting in higher detection accuracy while using fewer resources. They help reduce the model size and decrease inference time, making them an ideal choice for edge application devices. Finally, when working with hybrid systems that include a central analysis component, scalable data processing tools like Apache Spark are efficient. These platforms are designed for distributed data processing and can easily handle the large-scale data aggregation and analysis required for tasks like network-wide anomaly detection. This approach can truly complement the federated learning process, ensuring everything operates more smoothly.

3.2. Strengthening Federated Learning with Privacy-Enhancing Technologies (PETs).

To reduce the potential privacy risks associated with shared model updates in the FL workflow, various Privacy-Enhancing Technologies (PETs) can be employed. The most extensively used is Differential Privacy (DP). It uses a mathematical model but offers information survivability and privacy guarantees. It accomplishes this by adding a certain amount of statistically generated noise to the client-centered model before it is sent to the server. This noise obscures the influence of individual data points, making it difficult for data aggregators or potential attackers to access or infer specific information. Notably, DP relies on Gasper and analyses, who stated that “in order to guarantee privacy, a trade-off must be made between the amount of noise added, and the accuracy of the model.

Secure aggregation protocols are used to protect against a compromised or untrustworthy central server. These cryptographic techniques, such as Secure Multi-Party Computation (SMC) and Homomorphic Encryption (HE), allow the server to compute the sum or average of all client updates without decrypting or viewing any individual update. The aggregation occurs on encrypted data, with only the combined result shared. This ensures that even the training process coordinator cannot access the individual contributions from each client. Additional gradient-based defenses can further prevent information leakage. Methods like gradient clipping, which limits the maximum size of any gradient update, and gradient compression, which sparsifies the update vector, can reduce the amount of detailed information an attacker can extract from each update transmission.

3.3. Building Robust Defenses Against Adversarial Manipulation

The main method to defend against direct adversarial attacks on an AI model is through adversarial training. This involves adding adversarial examples, designed to fool the model, to the original training dataset. As a result, the model learns more robust and generalizable features, making its decision boundary more tolerant to small changes during inference. This approach enhances the model's resistance to known attack types. Another layer of defense includes input pre-processing and sanitization. These methods are applied to the data before it is used in the model. Techniques like adding random noise or smoothing filters to the input might disrupt the careful arrangement of an adversarial perturbation, causing it to fail in its purpose and allowing the model to classify the input correctly.

Finally, more advanced model-based defenses aim to alter a model's internal attributes to increase its resistance to external attacks. Some gradient masking techniques attempt to hide or distort model gradients, making it harder for attackers to design adversarial samples. Another method, known as defensive distillation, trains a second, "distilled" model using the soft probability outputs of a larger, initial model. This approach is designed to produce a smoother decision surface, reducing the model's sensitivity to small changes often exploited in adversarial attacks.

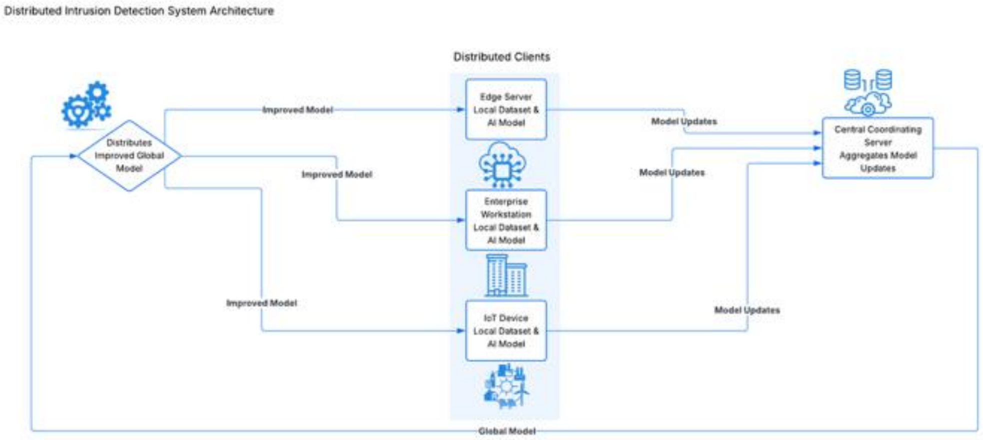
Distributed Intrusion Detection System Architecture
Distributed Clients
Distributes Improved Global Model
Improved Model
Improved Model
Improved Model
Edge Server Local Dataset & AI Model
Enterprise Workstation Local Dataset & AI Model
IoT Device Local Dataset & AI Model
Model Updates
Model Updates
Model Updates
Central Coordinating Server Aggregates Model Updates
Global Model

Fig 2: High-Level Architecture of the AI-Powered Distributed IDPS

3.4. Putting Performance into Practice: Insights from Case Studies

The practical benefits of having distributed AI-enabled IDPS systems are increasingly proven through various experiments and case studies across different domains. The value of these systems is often measured using standard evaluation metrics derived from a confusion matrix. This includes accuracy (the percentage of correct classifications out of all predictions), precision (the proportion of positive predictions that are true positives), recall (the proportion of actual attacks correctly identified), and the F1 score (the harmonic mean of precision and recall, providing a balanced measure of performance). Our analysis of results from major application areas shows notable performance improvements and effective handling of the challenges we previously mentioned.

Within the Internet of Things (IoT) realm, where devices are numerous and resources are limited, using FL-based IDPS proves to be very beneficial. Studies with realistic IoT datasets, including CICIoT2023 and ToN_IoT, show that distributed models can achieve detection accuracies above 98% and demonstrate high precision and recall, all while running on low-powered devices like the Raspberry Pi. A significant breakthrough in this field is the ability to maintain extremely high detection rates despite the constraints of low inference latency and small model sizes. This confirms that edge deployment is practical in real-world situations.

The Industrial Control Systems (ICS) and Industrial IoT (IIoT) domains emphasize security heavily, and data is often highly heterogeneous and siloed. Case studies in this area show how customized FL models can be effective. By allowing the global model to train, then fine-tuning on a local level, each node can adapt to the specific data distribution and operational routines of its industrial machines. This approach has been shown to reach over 95% accuracy, with some cases exceeding the performance of centralized models trained on non-IID (non-independent and identically distributed) data. It provides a strong solution to the data heterogeneity challenge in critical infrastructure environments.

Many benefits of large-scale enterprise and cloud environments include increased agility and reduced operational costs. Reports indicate that using distributed AI to identify and address threats can decrease the success rate of breaches by up to 30%. This improvement stems from better detection of zero-day threats and significantly fewer false positives compared to traditional systems. Additionally, by automating root-cause analysis and threat correlation, these systems greatly improve the Mean Time to Resolve (MTTR) for security incidents, allowing security operations teams to better manage complex systems and technologies.

Table 1: Summarizes Key Findings from Domains

Case Study / Domain	Architecture / AI Model	Key Performance Metrics & Improvements	Challenges Addressed
IoT/Smart Home Networks	FL CNN (convolutional neural networks) Models	Accuracy: ~98%; Precision/Recall/F1 >95%. Low latency and compact model size on edge hardware.	Privacy, Resource Constraints, Scalability
Industrial Control Systems (IIoT)	Personalized FL with CNN+GRU (gated recurrent units)	Accuracy: >95%, outperforming centralized models. F1-Score: ~0.94. Effective on non-IID data.	Data Heterogeneity (non-IID), Privacy, Critical Infrastructure Security
Enterprise/Cloud Environments	Distributed AI with Ensemble & DL Methods	Empirical evidence of up to 30% reduction in successful breaches. Significant decrease in false positives. Faster incident resolution (MTTR).	Scalability, Zero-Day Threat Detection, Operational Efficiency
General Cybersecurity (FL Benchmarks)	FL (FedAvg, FedProx) on UNSW-NB15, CICIDS2017	Accuracy >99% in FL settings, comparable to centralized performance. Detection accuracies >90% with privacy loss <5%.	Privacy-Preserving Collaboration, Data Silos

4. Conclusion

The growth of AI in distributed systems is arguably the most important development and a key part of intrusion detection and prevention systems. Systems built on frameworks like Federated Learning offer a strong solution to the limitations of traditional, centralized, and signature-based security, providing a path toward scalable, privacy-preserving, and adaptive security defenses. These systems, which can operate in a centralized manner, are highly scalable, adaptable to user needs, and, most importantly, preserve privacy while providing dynamic, situation-dependent defenses. Evidence demonstrates the potential for very high detection rates for both known and unknown threats across diverse settings, from resource-constrained IoT networks to critical industrial control systems. However, switching to these new methods poses significant challenges. The seamless development of these systems depends on complex trade-offs between performance, privacy, and security. These challenges are particularly intricate, relying heavily on computing limitations, the nearly invisible privacy leaks in federated model updates, and the vulnerability of AI systems to adversarial attacks. Progress in these systems is most likely to come from a shift in focus: moving from algorithmic tweaks to designing and implementing next-generation, layered hybrid defense systems that combine architectural innovations, covert cryptographic privacy methods, and AI model hardening frameworks.

References

[1] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Federated learning for cyber security: A comprehensive survey," IEEE Transactions on Neural Networks and Learning Systems [Online]. Available: https://arxiv.org/pdf/2108.00974

[2] Y. Li, Y. Chen, N. Li, and W. Lou, "A survey of privacy-preserving federated learning," ACM Computing Surveys, vol. 54, no. 6, pp. 1-36, 2021. [Online]. Available: https://scispace.com/pdf/a-survey-of-secure-computation-using-trusted-execution-2iwby4n5.pdf

[3] Belenguer, Aitor & Navaridas, Javier & Pascual Saiz, Jose Antonio. (2022). A review of Federated Learning in Intrusion Detection Systems for IoT. 10.48550/arXiv.2204.12443.

[4] Guembe, B., Misra, S., & Azeta, A. (2024). Privacy Issues, Attacks, Countermeasures and Open Problems in Federated Learning: A Survey. Applied Artificial Intelligence, 38(1). https://doi.org/10.1080/08839514.2024.2410504

[5] Tarrah R. Glass-Vanderlan, Michael D. Iannacone, Maria S. Vincent, Qian (Guenevere) Chen, and Robert A. Bridges. 2018. A Survey of Intrusion Detection Systems Leveraging Host Data. ACM Comput. Surv. 9, 4, Article 39 (March 2018), 39 pages. Available: https://www.osti.gov/servlets/purl/1965280

[6] Liang, Warren. (2023). Adversarial Attacks and Defense Mechanisms in AI-Based IDS for V2X. Available:

https://www.researchgate.net/publication/389089055_Adversarial_Attacks_and_Defense_Mechanisms_in_AI-Based_IDS_for_V2X

[7] Albulayhi, K., Smadi, A. A., Sheldon, F. T., & Abercrombie, R. K. (2021). IoT Intrusion Detection Taxonomy, Reference Architecture, and Analyses. Sensors, 21(19), 6432. https://doi.org/10.3390/s21196432

[8] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017. [Online]. Available: https://arxiv.org/abs/1602.05629

[9] Sowmya, T. & e a, Mary. (2023). A comprehensive review of AI based intrusion detection system. Measurement: Sensors. 28. 100827. 10.1016/j.measen.2023.100827.

[10] M. Z. Alom, T. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "A State-of-the-Art Survey on Deep Learning Theory and Architectures," Electronics, vol. 8, no. 3, p. 292, 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/3/292

[11] S. Agrawal, S. Sarkar, et al., "Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions," Computer Communications, vol. 160, pp. 415-425, 2021. [Online]. Available: https://www.semanticscholar.org/paper/Federated-Learning-for-Intrusion-Detection-System%3A-Agrawal-Sarkar/91b0acc50ff0b115ed4ce4010d0a471dac95d537

[12] L. N. R. Mudunuri, V. M. Aragani, and P. K. Maroju, "Enhancing Cybersecurity in Banking: Best Practices and Solutions for Securing the Digital Supply Chain," Journal of Computational Analysis and Applications, vol. 33, no. 8, pp. 929-936, Sep. 2024.